

# 유비쿼터스 컴퓨팅을 위한 지능적인 사용자 위치 이동 학습 및 예측

유지오, 김경중, 조성배

Department of Computer Science, Yonsei University  
134 Shinchon-Dong, Sudaemoon-Ku, Seoul, 120-749, Korea

Tel: +82-2-2123-3877, Fax: +82-2-365-2579, E-mail: {taiji391, uribyul}@sclab.yonsei.ac.kr, sbcho@cs.yonsei.ac.kr

연세대학교 공과대학 컴퓨터학과  
서울 서대문구 신촌동 134, 120-749

Tel: +82-2-2123-3877, Fax: +82-2-365-2579, E-mail: {taiji391, uribyul}@sclab.yonsei.ac.kr, sbcho@cs.yonsei.ac.kr

## Abstract

사용자의 지리적 위치에 따른 서비스를 제공하는 위치기반서비스는 유비쿼터스 컴퓨팅의 중요한 응용으로 여러 위치 감지기술과 다양한 시험 및 상용 서비스들이 개발되어 왔다. 하지만 기존의 위치기반서비스는 단순히 위치와 서비스를 정적으로 연결하는 기법에 그치고 있어 서비스의 유연성이 떨어지는 한계가 있다. 이를 개선하기 위해 위치 정보로부터 고수준 정보를 추론하여 보다 지능적인 서비스를 제공하려는 연구들이 이루어지고 있다. 본 논문에서는 사용자의 위치이동 데이터를 학습하여 미래의 위치 이동 경로를 예측하는 기법을 제안한다. GPS(Global Positioning System)를 사용하여 수집된 시퀀스 데이터를 시퀀스 데이터 처리에 특화된 RSOM(Recurrent Self Organizing Map)을 사용하여 클러스터링하고 이를 마르코브 모델을 사용하여 학습하여 각 위치 이동 패턴 모델을 구축한다. 현재의 위치이동 패턴을 구축된 각 이동패턴 모델들과 비교하여 가장 유사한 위치 이동패턴으로 미래의 사용자 이동을 예측한다. 제안한 위치이동 예측 기법을 평가하기 위해 실제 대학생의 생활을 기반으로 하여 GPS 데이터를 대학 캠퍼스 상에서 수집하고 이를 이용하여 제안한 방법의 학습 및 예측 성능을 평가한다. 그 결과 제안한 방법을 사용하여 사용자의 미래의 위치이동경로를 예측하는 것이 가능하고 불확실한 상황에서도 유연하게 예측을 수행함을 확인하였다.

## Keywords:

유비쿼터스 컴퓨팅; 위치기반 서비스; SOM(Self Organizing Map); 마르코브 모델

## 1. 서론

최근 모바일 컴퓨팅의 비약적인 발전은 고성능의 휴대용 기기와 빠른 속도의 무선 통신망을 가능하게 하였다. 이런 변화는 언제 어디서나 다양한 정보에 접근, 컴퓨팅 자원을 활용할 수 있는 유비쿼터스 컴퓨팅 기술의 개발로 이어지고 있다. 유비쿼터스 컴퓨팅은 우리의 일상을 한 단계 향상 시켜줄 수 있는 미래 지향적 기술 중 하나로 주목 받으며 활발히 연구되고 있다.

특히 저수준의 센서를 이용하여 사용자의 상태를 예측하기 위한 사용자 컨텍스트 모델링 방법에 관한 연구가 활발하다. 모바일 기기의 사용 목적이 일반적으로 사용자의 상황이나 위치, 시간 등과 밀접한 관련을 가지고 있는 경우가 많다. 그 중 위치 정보는 다른 컨텍스트에 비해 쉽게 얻을 수 있어 매우 광범위하게 사용되고 있는 정보이며 이를 활용한 연구가 활발히 진행되고 있다.

본 논문에서는 사용자 위치 이동패턴을 학습하여 다음에 움직일 위치를 추론하는 기법과 이를 실제 서비스 제공에 활용하는 방법을 제안한다. 사용자의 위치를 측정하기 위해 GPS(Global Positioning System) 신호를 사용한다. GPS 신호는 실외에서만 사용이 가능하고 10m의 오차 범위를 가지지만, 별다른 추가 알고리즘 없이 GPS 수신기만 있으면 쉽게 데이터를 얻을 수 있고, PDA(Personal Data Assistant) 등에 추가로 장착할 수 있는 상용 제품들이 많아서 쉽게 활용 가능하다는 장점이 있다. 수집된 사용자의 위치 정보들은 RSOM(Recurrent Self Organizing Map)을 통해 클러스터링되어 여러 개의 유사한 시퀀스 별로 마르코브 모델을 사용해 학습하여 사용자의 위치 이동 패턴을 모델링한다.

제안한 방법의 유용성을 보이기 위해 GPS 센서를 장착한 PDA에 대학생들을 대상으로 제안하는 방법을 적용한 비서 프로그램을 구현, 설치한다. 또한 실제

GPS 데이터를 바탕으로 학습을 수행하고 예측 성능을 측정한다.

### 1.1. 관련 연구

위치 정보의 활용에 관한 연구는 크게 정확한 위치 추적을 위한 방법에 대한 연구와 위치정보를 기초로 다른 정보를 추론하여 활용하는 연구로 나누어 볼 수 있다. 위치 추적에 사용되는 방법으로는 GPS 위성 신호, 무선 네트워크 신호, 초음파 신호, 카메라, 압력 센서 등 다양한 방법이 사용되어 왔으며 이러한 여러 가지 방법을 결합하여 더 정확한 추적을 가능하게 하려는 연구가 진행 중이다[1,2,3,4].

그러나 단순히 위치 정보만으로 새로운 지능형 서비스를 수행 하기에는 한계가 있기 때문에 다른 부가적인 정보를 얻는 연구도 활발히 진행되고 있다. D. Patterson 등은 GPS 신호로 사용자 위치를 추적하고 베이지안 네트워크로 GPS 신호를 모델링하여 현재 사용자가 이용하고 있는 이동 수단을 예측하는 방법을 제안하였다[5]. 또한 D. Ashbrook 등은 GPS 데이터를 k-means 알고리즘을 사용해 클러스터링하여 사용자가 오래 머무르는 유효한 위치를 찾아내고 이 위치간의 이동을 마르코브 모델을 사용하여 모델링하여 사용자 위치를 예측하도록 하였다[6]. 하지만 사용자가 머무르는 위치만을 사용하기 때문에 이동 중 경로에 따른 예측은 불가능하다. 또한 마르코브 모델은 매우 결정적(deterministic)이기 때문에 예측하는데 있어서 유연성이 떨어지는 단점이 있다.

## 2. 사용자 위치 이동 학습 및 예측

본 논문에서 제안하는 사용자 위치 이동 학습 방법은 세 단계로 이루어져 있다. 먼저 GPS 센서로부터 사용자의 위치 정보가 수집된다. 수집된 GPS 데이터는 SOM(자기 조직 신경망, Self Organizing Map)을 통해 학습된다. 학습된 SOM에서 비슷한 이동 패턴들은 출력 층의 같거나 가까운 노드에 사상되기 때문에 출력 층의 노드(연결 강도 벡터, 뉴런)는 사용자의 위치 이동 패턴을 의미하게 된다. 출력 층의 같은 뉴런에 사상된 GPS 데이터들은 마르코브 모델을 통해 하나의 지역 모델로 학습된다. 이 때 마르코브 모델은 SOM에서 GPS 데이터에 따라 변화하는 최적합 노드의 시퀀스 데이터를 바탕으로 학습된다.

사용자 위치 예측도 세 단계로 이루어져 있다. 먼저 현재 관측된 GPS 데이터가 SOM에 입력되어 최적합 노드가 출력된다. GPS 데이터가 입력됨에 따라 변화하는 최적합 노드의 시퀀스는 각 지역 모델에 의해 평가되는데, 어느 한 지역 모델이 다른 모델에 비해 일정 수준 이상 높은 확률로 평가되면 그 지역 모델에 할당된 서비스를 실행한다. 제안한 방법의 개요는 그림1과 같다.

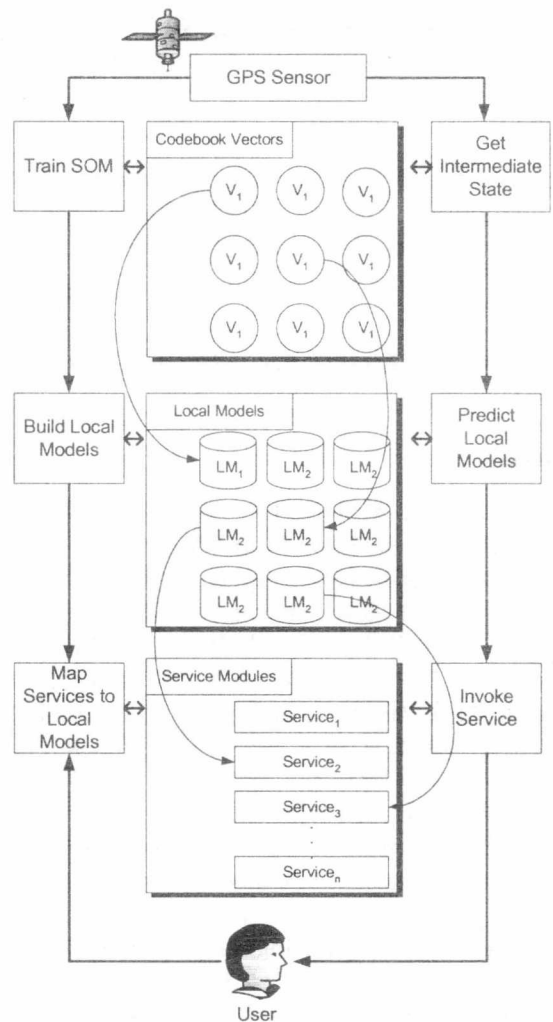


그림 1 - 제안한 방법의 개요

### 2.1. 사용자 이동 패턴 학습

사람이 일상 생활에서 이동하는 경로의 패턴은 너무 많기 때문에 이런 이동 패턴을 한번에 학습하는 것은 불가능하다. 그러므로 수집된 GPS 데이터를 클러스터링을 통해 작은 지역 문제로 나누어 처리하는 것이 요구된다. 제안한 방법에서 이러한 문제를 해결하기 위해 RSOM을 사용한다.

SOM은 대표적인 비교사 학습 신경 회로망으로 클러스터링이나 벡터 양자화 문제(Vector Quantization Problem)를 해결하는데 많이 사용되어 왔다. RSOM은 순차적인 데이터 처리에 적합하도록 변형된 것으로 기본적으로 일반적인 SOM이 가지고 있는 특징들을 모두 가지고 있다[7]. SOM은 비교사 학습을 하기 때문에 사전 지식이 없이도 유사한 이동 패턴들의 집합을 발견할 수 있어 사용자의 간섭 없이 일상 이동 패턴을 클러스터링 할 수 있는 장점이 있다. 또한 GPS 데이터의 순차적인 입력에 따른 최적합 노드의 변화 과정은 사용자 이동 패턴의 상태 변화로 볼 수 있다. 그러므로 GPS 데이터 대신 최적합 노드의 변화 과정을 학습할 경우 문제의 복잡도를 줄일 수 있



는 장점이 있다.

일반적인 SOM의 학습 알고리즘은 다음과 같다[8]. 여기서  $b(x)$  는 입력된 데이터에 따른 최정합 노드이고,  $\lambda$  는 이웃 반경 함수,  $\eta$  는 학습률 함수이다.

1. 연결 강도 벡터  $w_j(n)$  을 초기화 한다( $n=0$ ).
2. 입력된 데이터와 제일 가까운 노드를 찾는다.  

$$b(x) = \arg \min_j \|x(n) - w_j(n)\| \quad (1)$$
3. 각 연결 강도 벡터를 갱신한다.  

$$w_j(n+1) = w_j(n) + \eta(n)\lambda(b(x), j)(x(n) - w_j(n)) \quad (2)$$
4. 종료 조건이 만족될 때까지 2, 3의 과정을 반복한다.

RSOM은 순차적인 데이터를 처리하기 위해 일반적인 SOM과 다음과 같은 차이점을 가지고 있다. RSOM은 지난 연속적인 입력 벡터들로부터 발생한 오차를 누적하고, 이를 식(3)과 같이 현재의 오차 방정식에 적용한다.

$$y_i(n) = (1-\alpha)y_i(n-1) + \alpha(x(n) - w_i(n)) \quad (3)$$

여기서  $\alpha$  는 합성을 위한 가중치이고,  $y_i(n)$  은  $n$  번째 단계의 누적 오차 벡터,  $x(n)$  은  $n$  번째 단계의 입력 벡터이다. 또한 최정합 노드를 판단하는 방법과 그에 따라 가중치를 조정하는 방법도 일반적인 SOM과 조금 다르다.  $n$  번째 단계의 최정합 노드  $b(n)$  은 식(4)와 같이 구한다.

$$b(n) = \arg \min_i \|y_i(n)\| \quad (4)$$

$n$  번째 단계의  $i$  번째 가중치  $w_i(n)$  은 식(5)와 같이 갱신한다.

$$w_i(n+1) = w_i(n) + \eta(n)\lambda(b(n), i)y_i(n) \quad (5)$$

이와 같이 누적 오차 벡터를 가중치 조절에 반영해서 이전 시퀀스의 상태 정보를 고려할 수 있게 된다. 누적 오차 벡터는 각 입력 시퀀스의 학습이 끝날 때 0으로 초기화된 후 다음 입력 시퀀스의 학습에 사용된다. 그림 2는 RSOM의 학습 순서도를 보여준다.

본 논문에서 다루는 문제에 이 알고리즘을 적용시켜보면,  $x(n)$  은 시간이  $n$ 에 관찰된 GPS 데이터로 사용자의 위도와 경도를 벡터로 표현한 것이다. GPS 데이터는 사용자의 움직임에 상관없이 1초에 1~2회 간격으로 측정되기 때문에 의미 없는 데이터가 포함될 수 있다. 이를 해결하기 위해 GPS 데이터를 그대로 쓰지 않고, 사용자가 일정 범위 이상 움직였을 때의 GPS 데이터를 필터링하여 학습과 예측에 사용한다. 각 GPS 데이터 시퀀스  $SEQ_k = \{x(0), x(1), \dots, x(N-1)\}$ 는 RSOM에서 가장 마지막 최정합 노드에 따라 클러스터링된다.  $i$  번째 노드에 할당된 지역 모델의 학습 데이터는 해당 노드에 클러스터링된 데이터 입력에 따른 최정합 노드의 변화 시퀀스의 집합  $D_i = \{M_1, M_2, \dots, M_K\}$ 가 사용된다. 이 때  $M_i = \{b(0), b(1), \dots, b(N-1)\}$ 은 한 GPS 데이터 시퀀스에 따른 최

정합 노드의 변화 시퀀스이다.

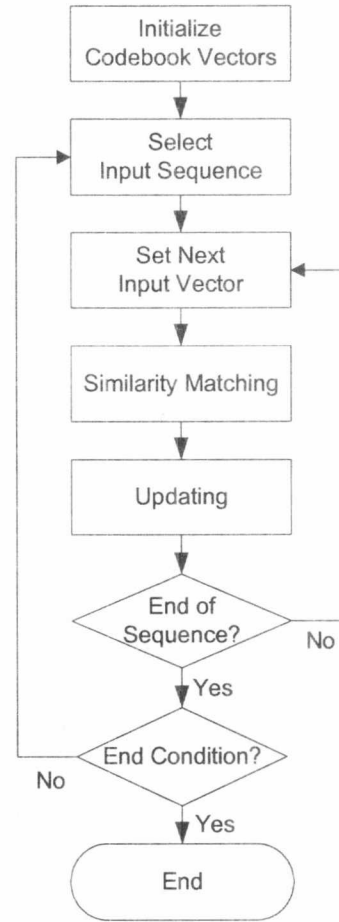


그림 2 - RSOM의 학습 순서도

## 2.2. 지역 모델 학습

RSOM을 이용해 지역 데이터 집합으로 클러스터링된 사용자 이동 패턴은 마르코브 모델에 의하여 학습된다. 출력 노드수 만큼의 지역 모델이 마르코브 모델을 이용하여 만들어 진다. 마르코브 모델은 초기 상태 분포  $Q$ 와 상태 전이 확률 분포 행렬  $P$ 로 이루어진다. 사용자의 이동에 따른 최정합 노드의 변화과정이 입력으로 사용되므로 마르코브 모델 상태수  $n$ 은 RSOM의 출력 노드 수만큼의 상태를 가진다. 시간  $t$ 의 상태  $i$ 에서 시간  $t+1$ 의 상태  $j$ 로 전이될 확률을  $p_{ij}$ 라고 할 때, 상태 전이 확률 행렬  $P$ 는 식(6)과 같이 정의된다. 한 상태에서 다른 상태로 전이될 확률을 모두 합하면 1이 된다.

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} \quad (6)$$

$$\sum_{j=1}^n p_{ij} = 1 \quad (7)$$

그리고 시간 0에서 상태  $i$ 가 될 확률을  $q_i$ 라 할 때 초기 확률 분포  $Q$ 는 다음과 같이 1차원 행렬로 표현한다.

$$Q = [q_1 \quad q_2 \quad \cdots \quad q_n] \quad (8)$$

마르코브 모델의 학습은 데이터로부터 상태 전이 확률 분포 행렬  $P$ 와 초기 확률 분포  $Q$ 를 얻어내는 과정이라고 할 수 있다. 시간 0부터  $N-1$ 까지 상태 변화 시퀀스  $M_i = \{b(0), b(1), \dots, b(N-1)\}$ 가 학습 데이터로 주어졌다고 할 때, 다음과 같은 식(9)와 식(10)에 의해  $P$ 와  $Q$ 를 얻어낼 수 있다.

$$p_{ij} = \frac{N_{ij}}{N_i} \quad (9)$$

$N_{ij}$  : 시퀀스 중 상태  $i$ 에서  $j$ 로의 전이 회수

$N_i$  : 시퀀스 중 상태  $i$ 가 나타난 회수

$$q_i = \frac{N_i}{N} \quad (10)$$

$N_i$  : 시퀀스 중 상태  $i$ 가 나타난 회수

$N$  : 모든 관찰 시퀀스의 수

이와 같이 RSOM에서의 최적합 노드의 시퀀스 데이터를 사용하여 학습을 했기 때문에 마르코브 모델의 상태  $i$ 는 RSOM의 노드  $i$ 를 의미하게 된다. 마르코브 모델인 지역 모델  $LM_i$ 에서 최적합 노드의 시퀀스  $b(0), b(1), \dots, b(T)$ 가 발생할 확률은 다음과 같은 식(11)에 의해 계산된다.

$$P(b(0), b(1), \dots, b(N) | LM_i) = q_{b(0)} \prod_{t=2}^T p_{b(t-1)b(t)} \quad (11)$$

그림 3은 RSOM과 마르코브 모델이 어떻게 연동되는지를 보여준다.

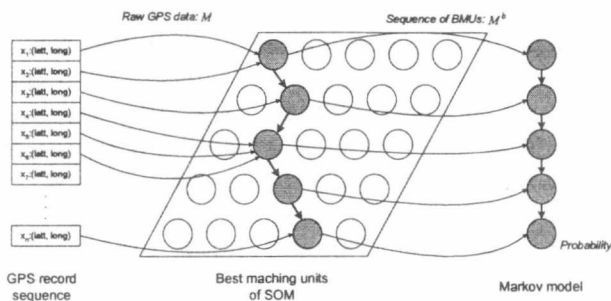


그림 3 - RSOM과 마르코브 모델의 연동

### 2.3. 미래의 움직임 예측

이렇게 구축된 지역 모델을 사용하여 사용자 위치 이동을 예측한다(그림 4). 사용자의 현재 움직임이 GPS 데이터를 통해 RSOM에 입력되면 그에 해당하는 최적합 노드가 출력된다. 현재까지 출력된 노드 시퀀스는 각 지역 모델에 의해 평가되며 다른 지역 모델에 비해 특히 높은 확률을 보이는 지역 모델이 있을 경우 즉시 선택되어 사용자의 움직임을 예측한다.

다. 현재 입력된 시퀀스를  $M_i = \{b(0), b(1), \dots, b(N-1)\}$ 라 할 때, 지역 모델  $LM_j$ 에 대한 *significance*는 다음 식(11)과 같이 계산한다( $n$ 은 지역 모델  $LM$ 의 총 수). 이 값이 식(12)와 같이 임계치를 넘을 경우 해당하는 모델이 사용자의 움직임을 예측하는 모델로 선택된다.

$$\text{significance}(LM_i) = \frac{P(M | LM_i) - \sum_{j=1, j \neq i}^n \frac{P(M | LM_j)}{n-1}}{n-1} \quad (11)$$

$$\begin{cases} LM_i \text{ select} & \text{significance}(LM_i) > \text{threshold} \\ \text{none} & \text{significance}(LM_i) \leq \text{threshold} \end{cases} \quad (12)$$

이와 같이 지역 모델의 *significance*는 자신을 제외한 다른 모델들의 확률 값의 평균과 자신의 확률 값의 차이를 의미하고, 이 차이가 미리 정의된 임계치보다 커지게 되면 사용자의 움직임이 해당 모델이 학습한 시퀀스 데이터와 유사하다고 판단할 수 있다. 임계치가 작으면 사용자 움직임에 대한 예측이 빨라질 수 있지만 정확한 예측을 보장받기 어렵다. 만약 임계치가 크면 예측은 늦게 일어나겠지만 잘못 판단하는 경우는 적어질 것이다.

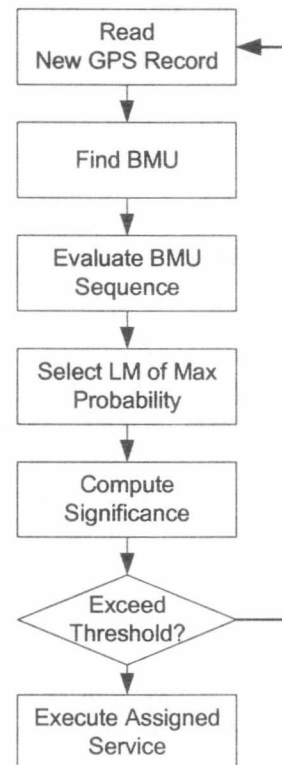


그림 4 - 사용자 이동 예측 순서도

BMU : Best Matching Unit, 최적합 노드,  
LM : Local Model, 지역 모델.

## 3. 구현

위치 이동 예측은 많은 응용 프로그램에서 사용될



수 있다. Ashbrook과 Starner는 사용자가 혼자이거나 여럿일 환경에서 위치 이동 예측을 사용 가능한 응용 프로그램을 소개한 바 있다[6]. 본 논문에서는 제한한 사용자 위치 이동 예측 방법을 사용한 사용자 비서 프로그램의 프로토타입을 구현하였다. 프로그램의 주 사용자는 대학생으로 설정하였고, 사용자는 GPS 수신기가 달린 PDA를 들고 다니며 서비스를 받게 된다. 프로그램은 사용자의 캠퍼스 내 위치 이동을 예측하여 그에 알맞은 서비스를 제공해준다.

### 3.1. 하드웨어

마이크로소프트 Pocket PC 기반의 HP iPAQ PDA에 사용자 비서 프로그램을 구현하였다. 또한 GPS 수신기를 추가로 장착하여 사용자의 위치를 측정한다. 사용자는 PDA에 내장된 무선 랜을 통해 캠퍼스 내 어디에서나 월드 와이드 웹에 접속할 수 있다. 그림 5는 GPS 수신기가 장착된 PDA에 비서 프로그램이 구현된 모습이다.

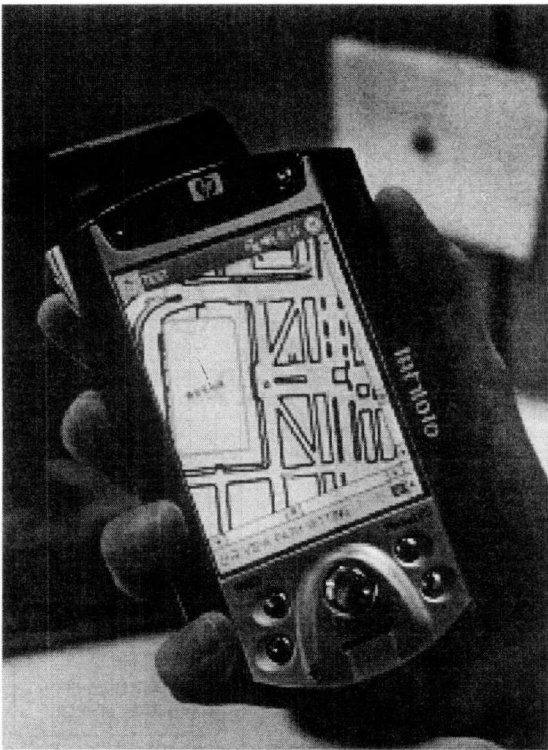


그림 5 - GPS 수신기가 장착된 PDA

### 3.2. 소프트웨어

PDA에서 동작하는 프로그램을 제작하기 위해서 Embedded Visual C++를 사용하였다. 그림 6(a)는 RSOM을 통해 분류된 사용자 이동 패턴을 보여주는 사용자 인터페이스이다. 2차원 격자 안의 버튼들은 RSOM의 출력 노드를 의미하고, 활성화된 버튼은 유의한 이동 패턴들을 나타낸다. 사용자 이동 경로에 따라 클러스터링된 출력 노드만 활성화되어 자신의 이동 패턴이 몇 가지 집단으로 구분되었는지 확인

할 수 있다. 또 사용자는 활성화된 버튼을 선택하여 실제 할당된 사용자 이동 경로를 지도 상으로 확인할 수 있다. 이동 경로를 확인한 후 각 이동 경로에 맞는 서비스를 선택할 수 있다. RSOM의 인접한 출력 노드에 할당된 이동 경로들은 서로 유사하기 때문에 인접 노드에 할당된 서비스를 참고해서 쉽게 서비스를 선택할 수 있다. 그림 6(b)는 GPS 데이터를 지도상에 보여주고 그에 해당하는 서비스를 선택하는 장면이다.

U\_Agent

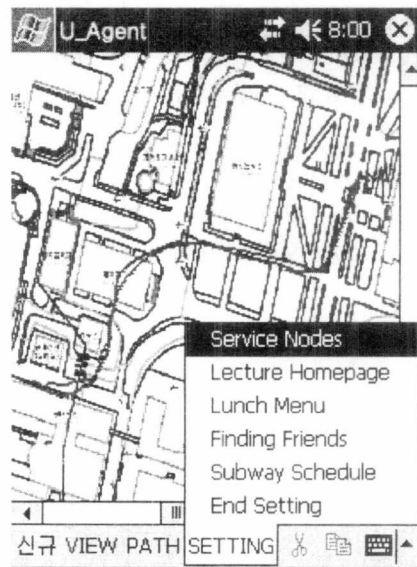
5:28

<b>1,1</b>	<b>1,2</b>	1,3	1,4	<b>1,5</b>	1,6	<b>1,7</b>
2,1	2,2	2,3	2,4	2,5	2,6	2,7
3,1	3,2	3,3	3,4	3,5	3,6	<b>3,7</b>
4,1	4,2	4,3	<b>4,4</b>	4,5	4,6	4,7
<b>5,1</b>	5,2	5,3	<b>5,4</b>	5,5	5,6	5,7
6,1	6,2	6,3	<b>6,4</b>	6,5	6,6	<b>6,7</b>
<b>7,1</b>	7,2	7,3	7,4	7,5	7,6	7,7

Service

ADD / REMOVE

(a)



(b)

그림 6 - 사용자 인터페이스

또한 사용자 이동 경로에 따른 서비스를 구현하였다. 그림 7은 제공 가능한 서비스들을 보여주고 있다. 이 서비스들은 이동 경로를 참조하여 사용자에게 유용한 서비스를 제공할 수 있다. 예를 들어 특정 건물에 접근하는 이동 경로에 강의 홈페이지 서비스를 할당하여 사용자가 수업 장소에 도착하기 전에 그

날 수업의 주제를 미리 알 수 있도록 해주거나, 학생 식당에 도착하기 전에 미리 점심 메뉴를 보여주거나, 혹은 집으로 돌아가기 위해 지하철 역으로 향하면 지하철 노선과 운행 스케줄을 보여줄 수 있다.

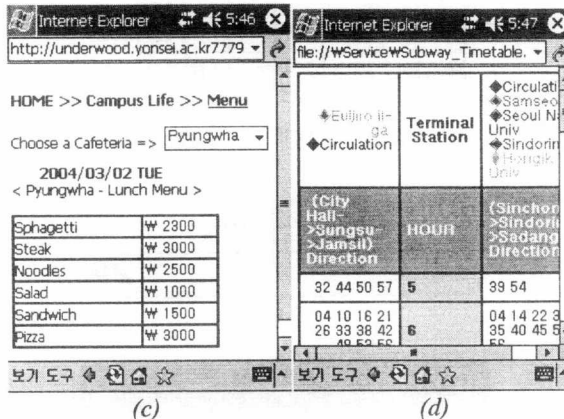
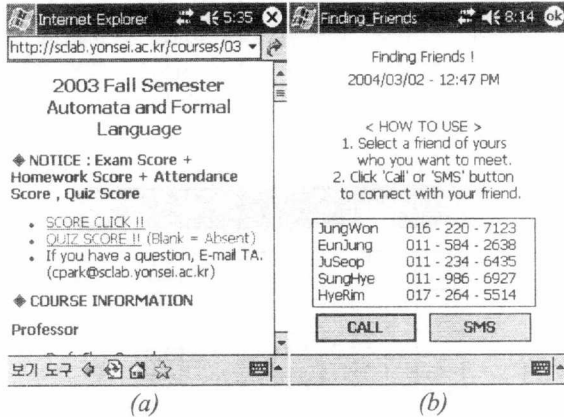


그림 7- 제공되는 서비스

- (a) 강의 홈페이지 열람,
- (b) 친구 찾기,
- (c) 점심 메뉴 표시,
- (d) 지하철 운행 스케줄.

## 4. 실험

### 4.1. 실험 데이터

제안한 방법의 유용성을 보이기 위해 대학생의 캠퍼스 생활을 바탕으로 실제 GPS 데이터를 수집하였다. 표 1과 같이 9개 장소에 따라 수업 듣기, 점심 식사하기, 공부하기, 동아리 활동 등 시나리오를 설정하고 이에 따라 10가지 종류의 경로 데이터를 수집하였다. 2명의 사용자가 미리 정의된 경로를 GPS 센서가 장착된 PDA를 들고 다니며 GPS 데이터를 수집하고, 수집된 데이터는 잘못 측정된 정보를 제거한 후 시작 장소와 도착 장소를 기준으로 분류하였다. 10가지 종류의 경로 데이터 당 6개의 케이스를 수집하여 총 60가지 이동 경로를 바탕으로 학습을 수행

하였다. 그림 8은 수집된 경로 데이터를 지도 상에 표시한 것이다. GPS의 기본적인 예러 때문에 수집된 데이터는 실제 이동 경로와 완전히 일치하는 것은 아니다.

표 1- 수집된 이동 경로

번호	이동 경로	설명
1	정문→제1공학관	수업
2	제1공학관→제2인문관	아르바이트
3	제2인문관→대강당	수업
4	대강당→사회과학관	수업
5	사회과학관→제3공학관	수업
6	제3공학관→학생회관	점심 식사
7	학생회관→제3공학관	수업
8	제3공학관→중앙도서관	공부
9	중앙도서관→제1인문관	동아리
10	제1인문관→정문	귀가

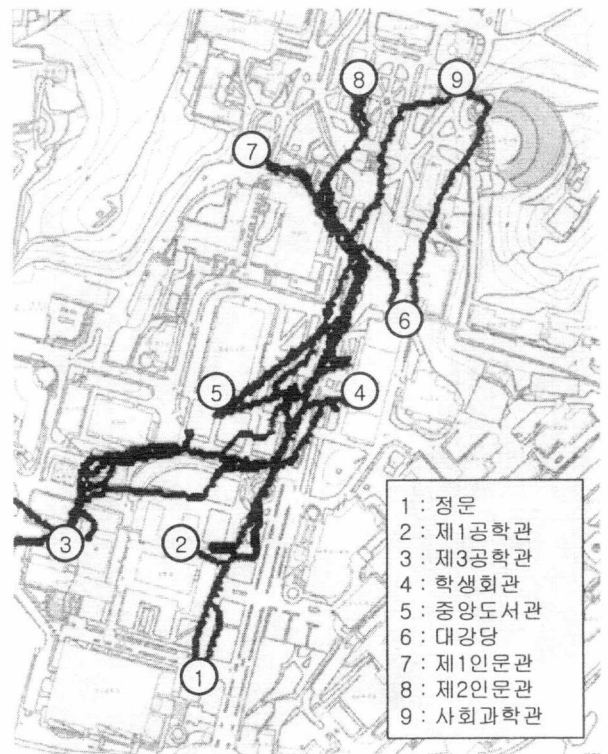


그림 8- 이동 경로 표시

### 4.2. 실험 결과

클러스터링: 먼저 RSOM의 클러스터링 정도를 평가하였다. RSOM의 학습 과정 이후 각 이동 경로들은 마지막 시퀀스 데이터의 최정합 노드에 할당되고, 이 때 각 출력 노드들은 이동 경로를 표상하며 10개의 이동 경로 중 하나로 할당된다. 출력층의 크기는 8 X 8, 초기 반경은 4, 초기 학습률은 0.03으로 설정하였다. 학습은 1000번 실행되고 종료되도록 하였다. 그림 9는 학습된 후 RSOM의 출력 노드에 할당된



경로들을 나타낸다.

각 원은 RSOM의 출력 노드를 의미하고, 회색 원은 클러스터링에 포함되지 않은 노드를 의미한다. 그림에서 보는 바와 같이 유사한 이동 경로는 인접한 노드에 위치하고 있음을 알 수 있다. 또한 클러스터링된 패턴은 크게 왼쪽 위의 그룹과 오른쪽 아래 그룹으로 나뉘고 있음을 알 수 있다. 왼쪽 위의 그룹에는 1, 5, 6, 7, 8, 10 번의 이동 경로가 할당되어 있고, 오른쪽 아래의 그룹에는 2, 3, 4, 9 번의 이동 경로가 할당되었다. 왼쪽 위의 그룹에 할당된 경로들은 도착지점이 제1공학관, 제3공학관, 정문, 학생회관, 중앙 도서관 등 모두 캠퍼스 지도 상의 남쪽 부분에 위치하고 있다. 반대로 오른쪽 아래의 그룹에 속한 경로들은 도착지점이 대강당, 제1인문관, 제2인문관, 사회과학관 등 모두 캠퍼스 지도 상의 북쪽 부분에 위치하고 있었다. 이는 SOM의 위상 보존에 의한 것으로 이와 같은 성질은 사용자가 이동 경로에 대해 파악하고 적절한 서비스를 쉽게 할당할 수 있도록 도와준다.

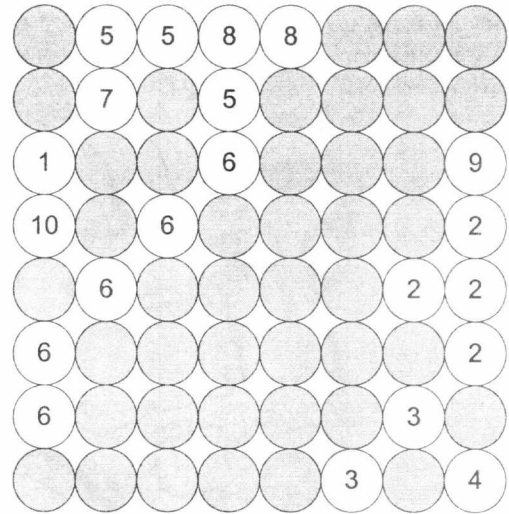


그림 9-RSOM의 출력 노드

예측: 그림 10은 시간에 따른 최정합 노드의 변화를 보여준다. 그림에서 보는 바와 같이 이동 패턴이 다르면 최정합 노드의 변화 시퀀스와 최종 최정합 노드가 달라지고 있음을 알 수 있다. 그러므로 현재의 최정합 노드의 변화 과정과 지난 변화 과정의 비교를 통해 사용자의 이동 경로를 예측할 수 있다고 할 수 있다.

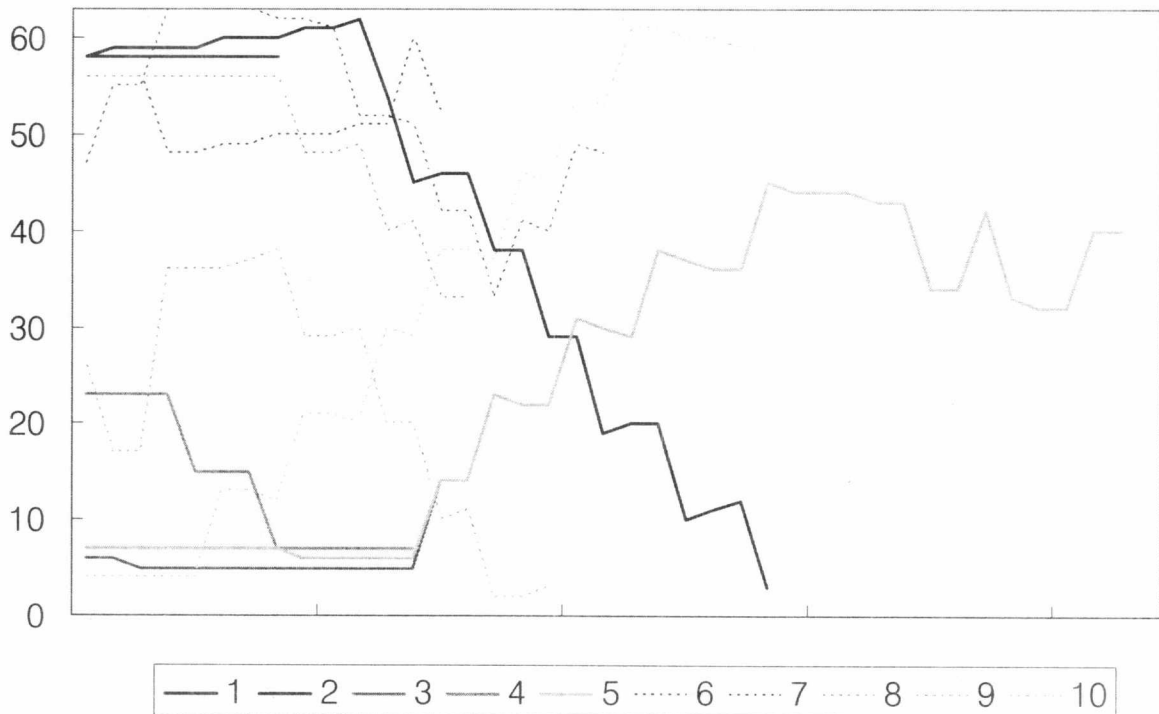


그림 10- 시간에 따른 최정합 노드의 변화

X축: 최정합 노드의 Index,

Y축: 시간,

각 선은 이동 경로를 의미함.



시간	지역 모델 번호(학습 데이터 번호)																	
	5(1)		5(2)		6(1)		6(2)		6(3)		6(4)		6(5)		8(1)		8(2)	
1	-1.6	1.6	-2.8	2.7	-0.6	0.4	-0.5	0.4	-1.2	1.0	-1.1	1.0	-0.8	0.6	-1.0	0.8	-0.3	0.1
2	-1.9	1.7	-3.0	2.9	-0.8	0.5	-0.6	0.4	-1.3	1.1	-1.3	1.1	-0.9	0.7	-1.2	1.0	-0.4	0.2
3	-2.0	1.9	-3.1	3.0	-0.8	0.6	-0.7	0.5	-1.5	1.3	-1.4	1.2	-1.0	0.8	-1.4	1.2	-0.5	0.3
4	-2.2	2.0	-3.3	3.1	-0.9	0.7	-0.8	0.6	-1.6	1.4	-1.6	1.4	-1.1	0.9	-1.6	1.4	-0.5	0.3
5	-2.3	2.1	-3.5	3.3	-1.1	0.8	-0.9	0.6	-1.8	1.6	-1.7	1.5	-1.2	1.0	-1.9	1.6	-0.6	0.4
6	-2.4	2.2	-3.6	3.4	-1.3	0.9	-1.0	0.7	-1.9	1.7	-1.9	1.7	-1.3	1.1	-2.1	1.9	-0.7	0.4
7	-2.6	2.3	-3.8	3.6	-1.3	1.0	-1.0	0.8	-2.1	1.8	-2.0	1.8	-1.4	1.2	-2.3	2.1	-0.8	0.5
8	-2.7	2.4	-3.9	3.7	-1.4	1.1	-1.1	0.9	-2.2	2.0	-2.2	1.9	-1.5	1.3	-2.5	2.3	-0.8	0.6
9	-4.8	4.2	-5.9	5.4	-3.6	3.1	-3.6	3.1	-4.2	3.7	-4.1	3.6	-3.8	3.3	-4.1	3.6	-3.5	3.0
10	-5.9	5.4	-6.9	6.4	-4.3	3.9	0	0	-4.5	4.0	0	0	-4.5	4.0	-4.8	4.3	-4.2	3.7
11	0	0	0	0	0	0	0	0	-5.8	5.8	0	0	-5.2	5.1	0	0	0	0
12	0	0	0	0	0	0	0	0	-6.5	6.5	0	0	-5.6	5.5	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	-6.7	6.7	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	-7.4	7.4	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	-8.1	8.1	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

그림 11-각 지역 모델의 확률과 significance 값의 변화  
 각 모델의 왼쪽 열은 확률, 오른쪽 열은 significance 값을 의미함.



그림 11은 제3공학관에서 학생회관으로 이동하는 이동 패턴 6을 따라 이동하면서 각 지역 모델이 나타내는 확률과 *significance*값을 나타낸 그림이다. 확률이 계속 0을 나타내는 지역 모델은 표시하지 않았다. 그림에서 회색으로 칠해진 칸은 각 시간대에서 최고 확률을 보이고 있는 지역 모델임을 의미한다. 사용자는 제3공학관에서 움직이기 시작하는데, 이동 초기에는 8(2)번 지역 모델이 가장 높은 확률을 보여주고 있다. 이는 제3공학관을 출발하여 중앙도서관으로 가는 이동 패턴 8이 이동 초기에는 이동 패턴 6과 비슷하기 때문이다. 그러나 *significance*값은 아직 작아서 5.5로 설정된 임계치를 넘지 못하고 있기 때문에 예측은 일어나지 않고 있다.

시간 11에서 사용자가 분기점을 통과하면서 사용자가 움직이는 방향이 더 이상 중앙도서관 쪽이 아니라는 점이 더 분명해진다. 그렇기 때문에 상태 변화에 있어서 8(2)번 지역 모델의 확률은 거의 0으로 떨어지고, 6(5)번 지역 모델의 확률이 가장 높은 확률이 된다. 시간 12에서 6(5)번 지역 모델에 대한 *significance*값이 임계치를 넘으면서 사용자가 이동 패턴 6과 비슷한 경로로 진행할 것으로 예측한다. 이에 따라 이동 패턴 6에 할당되어 있는 서비스가 있는지 조사하고 결국 6(2)에 할당되어있는 점심 메뉴 띄우기 서비스를 실행하게 된다.

사용한 데이터가 많지 않기 때문에 실험은 6-fold cross validation으로 진행되었다. 전체 데이터를 10개의 이동 패턴이 포함된 여섯 부분으로 나누고, 그 중 다섯 부분은 학습에 사용하고 나머지 한 부분은 예측하는데 사용하였다.

표 2는 예측 실험의 결과를 정리한 것이다. “Miss”는 계속 낮은 *significance*값이 출력되어 예측을 못한 경우의 수를 의미하고 “Error”는 예측은 했으나 예측한 값이 잘못되었을 경우를 의미한다. 각 테스트 데이터마다 10개의 이동 패턴을 실험한 결과 “Miss”와 “Error”인 경우가 0~1회만 나타나 좋은 성능을 보이고 있음을 알 수 있다. 대부분의 “Miss”는 이동 패턴 1에서 발생하였다. 이는 정문에서부터 제1공학관까지 가는 거리가 짧아 임계치를 넘길만한 충분한 *significance*값을 발생시킬 여유가 없었기 때문이다. 예측까지 걸리는 시간도 측정하였는데, 전체 패턴에 대해 사용자의 평균 이동 시간은 5분 25초였으며 평균 예측 시간은 1분 31초였다. 이동 패턴 1과 같이 이동 시간이 평균 예측 시간보다 짧은 경우 예측이 잘 되지 않을 수 있으므로 이동 거리에 맞게 각 지역 모델의 임계치를 조정해줌으로써 예측 시간을 조정할 수 있다.

표 2- 예측 실험 결과  
(총 10개의 패턴 중 여러 회수)

Fold	1	2	3	4	5	6	Avg.
Miss	1	0	0	1	1	1	0.67
Error	1	1	0	1	1	1	0.83

## 5. 결론

본 논문에서는 사용자의 위치 이동 경로를 학습하고 미래의 위치 이동을 예측하는 새로운 방법은 제안하였다. 사용자의 움직임은 GPS 데이터를 통해 측정되고, 이를 RSOM을 통해 유의한 이동 패턴을 찾아냄과 동시에 GPS 데이터를 클러스터링하여 작은 지역 문제로 나누어 적용 가능하도록 하였다. 발견된 유의한 이동 패턴은 마르코브 모델을 통해 지역 모델로 학습된다. 현재 이동 경로는 각 지역 모델이 평가하여 어느 한 지역 모델이 나머지 지역 모델보다 일정 수준 이상 높은 확률을 보이면 그 지역 모델을 학습하는데 사용한 이동 경로가 현재 사용자의 이동 경로임을 미리 예측할 수 있었다.

제안한 방법의 유용성을 보이기 위한 응용 프로그램으로 대학생의 실제 생활에 기반한 비서 프로그램을 구현하였다. 사용자는 GPS 센서가 장착된 PDA를 휴대하고 다니며 사용자의 현재 이동 경로에 따른 서비스를 사용자에게 제공할 수 있었다. 또한 RSOM의 위상 보존 특성 때문에 사용자가 쉽게 자신의 이동 패턴을 찾아보고 서비스를 할당할 수 있었다.

향후 연구로는 실제 대학 생활에서 발생할 수 있는 GPS 데이터를 좀더 방대하게 모아서 제안한 방법을 좀더 정확하게 분석하고 성능을 비교해볼 예정이다. 또한 대학 생활에서 유용할 것으로 예상되는 서비스를 더 개발하고 구현하여 실제 대학생들을 통해 사용자 테스트를 하고, 유용성을 검증할 필요도 있다.

## 감사의 글

본 연구는 21세기 프론티어 연구개발사업의 연구비 지원에 의해 수행되었습니다.

## 참고 문헌

- [1] J. Hightower and G. Borriello (2001). “Location Systems for Ubiquitous Computing,” *IEEE Computer*, Vol. 34, No. 8, pp. 57-66.
- [2] J. Hightower and V. Padmanabhan (2000). “RADER: An In-Building RF-based User Location and Tracking System,” *Proceedings of IEEE Infocom 2000*, pp. 775-784.
- [3] R.J. Orr and G.D. Abowd (2000). “The Smart Floor: A Mechanism for Natural User Identification and Tracking,” *Proceedings of Conference on Human Factors in Computing Systems*, pp. 275-276.
- [4] J. Krumm, S. Harris, B. Meyers, B. Brumitt, M. Hale, and S. Shafer (2000). “Multi-Camera Multi Person Tracking for Easy Living,” *Proceedings of the 3<sup>rd</sup> IEEE International Workshop Visual Surveillance*, pp. 3-10.
- [5] D. Patterson, L. Liao, D. Fox and H. Kautz (2003). “Inferring High-Level Behavior from Low-Level Sensors,” *Proceedings of the Fifth International*

*Conference on Ubiquitous Computing*, pp. 73-89.

- [6] D. Ashbrook and T. Starner (2002). "Learning Significant Locations and Predicting User Movement with GPS," *Proceedings of IEEE Sixth International Symposium on Wearable Computing*, pp. 101-108.
- [7] T. Koskela, M. Varsta, J. Heikkonen and K. Kaski (1998). "Temporal Sequence Processing using Recurrent SOM," *Proceedings of Second International Conference on Knowledge-Based Intelligent Engineering Systems*, Vol. 1, pp. 290-297.
- [8] T. Kohonen (1990). "The Self-organizing Map," *Proceedings of the IEEE*, Vol. 78, No. 9, pp.1464-1480